REMARKS

Reconsideration and allowance in view of the foregoing amendment and the following

remarks are respectfully requested. Claims 7, 15, 19 and 25 are amended without prejudice or

disclaimer. Claim 18 has been cancelled without prejudice or disclaimer.

Objection to the Drawings

The Office Action objects Figure 1 and requires that it should be designated by a legend

such as "Prior Art". Applicants provide herein a replacement sheet for Figure 1 and request

withdrawal of the rejection to Figure 1.

Objection to Claims 16 and 18

The Office Action objects claims 16 and 18 because of informalities. Applicants have

cancelled claim 18 thus rendering this claim rejection moot.

Rejection of Claims 1-35 Under 35 U.S.C. §102(b)

The Office Action rejects claims 1-35 under 35 U.S.C. §102(b) as being anticipated by

Gasper et al. (U.S. Patent No. 5,278,943) ("Gasper et al."). Applicants traverse this rejection and

respectfully submit that Gasper et al. fail to teach each limitation of the claims.

The Office Action on page 3 asserts that Gasper et al. teach the step of claim 1 of

receiving a user selection of a first text-to-speech (TTS) voice and a second TTS voice from a

plurality of TTS voices is taught at column 4, line 53 – column 5, line 37. For ease of discussion

this portion of the reference is reproduced as follows:

> "In the preferred embodiment, linear prediction coding (LPC) is utilized to encode the
> speech data derived from actual speech samples. Prior art methods of speech data
> representation typically utilize LPC to encode and store speech data. Short segments of
> sampled speech data (frames) comprising a substantial number of samples are converted
> to a linear filter model and a residual vocal tract excitation signal of the same length
> representing the airflow into the vocal tract. The airflow typically consists of fricative
> noise from the lungs and pulses from the glottis. For a 1/60 second (s) frame of sample
> data at 22 kHz containing 370 samples, the filter model is typically represented by 10 to
> 12 bytes of data and the residual excitation signal by another 370 bytes of data. It is
> known in the prior art that acceptable speech can be produced by reducing the residual

excitation signal to a few simple parameters (e.g., energy level, voice/unvoiced indicators) which can be represented in 1 or 2 bytes of data. During resynthesis, the excitation is modelled by a noise generator and a pulse generator and prosodic variation can be introduced into the stored speech data. This method is very compact, but the airflow modeling techniques utilized yield low quality, mechanical sounding speech due to the fact that they are artificially generated as discussed hereinabove.

One advantage of LPC representation over noncoded sampled data representation is a reduction of the storage requirements. In the example given above, 370 data samples were compressed to 12 to 14 bytes, a substantial savings. Another advantage is that because the pitch and energy level of the synthesized speech is dependent on the vocal tract excitation, conventional speech synthesizers can vary the pitch and energy level of the original data by varying the artificially generated excitation to the filter models. This technique has been used successfully in the prior art to produce acceptable synthetic speech.

A major limitation of the prior art technique to encode speech data using LPC described elsewhere in this specification is that much of the speaker-dependent information contained in the residual excitation signal has been discarded. The residual excitation signal contains information about the speaker's lungs and glottis which is amplified by the speaker's vocal tract and contributes greatly to the individuality and identification of the speaker's voice. In one preferred embodiment of the present invention, an enhanced LPC data representation is used which stores the residual excitation signal rather than generating it artificially. This technique retains all of the advantages of the prior art LPC representation while minimizing the loss of speaker-dependent information from the residual excitation signal."

This portion of the reference discusses how linear predictive coating (LPC) is utilized to encode the speech data derived from actual speech samples. There are a number of details regarding the advantages of LPC representation over non-coded samples. Applicants respectfully submit that as can be seen above, wherein this portion of the reference merely highlights the advantages of LPC encoding of speech data, they fail to teach anything regarding receiving a user selection of a first TTS voice and a second TTS voice from a plurality of TTS voices. It is almost as though the Examiner miscited an intended portion of Gasper et al. In any event, the cited portion fails to teach anything regarding this first limitation of claim 1 and accordingly, Applicants submit that this limitation is not taught in the reference.

Next, the Office Action asserts that column 4, line 53 – column 5, line 37 teach the step of "receiving at least one user-selected voice characteristic." Applicants also traverse that the

above-cited portion of Gasper et al. teach this particular limitation. Again, this portion of the reference is primarily focused on encoding of speech data derived from actual speech samples. There is no teaching of a system that receives at least one user-selected voice characteristic. Accordingly, Applicants submit that this feature is not taught in the reference.

Finally, claim 1 recites "generating a new TTS voice and the second TTS voice and according to the user selected voice characteristic." The Office Action asserts that this taught in column 5, lines 50-64 and column 6, lines 58-65. The portion cited in column 5 fails to teach anything regarding blending a first TTS voice and a second TTS voice. This portion of the reference does teach "prosodic variation is then added utilizing the defined prosodic rules." However, absent from this portion is any discussion from a first and second TTS voice. The portion cited in column 6 discusses how the voice animation system of the invention of Gasper et al. provides a "method of animation for mimicking an individual voice, creating new artificial voices or for combining the two." The Office Action on page 3 misstates the particular limitation of claim 1 as "generating a new TTS voice and the second TTS voice and according to the user selected voice characteristic."

Applicants note that this fundamentally differs from the limitation of claim 1 which requires "generating a new TTS voice by blending the first TTS voice and the second TTS voice and according to the user selected voice characteristic." Accordingly, the teaching in column 6 of combining the two [i.e., combining an individual voice and a created new artificial voice] differs from the particular limitation of claim 1.

Applicants also note that a more appropriate interpretation of this sentence in column 6 would be that the voice animation system of the invention of Gasper et al. can provide a method of animation for mimicking an individual voice, creating new artificial voices or for "combining the two" in the sense of a system that offers both functions. In other words, a system that

enables a user to mimic an individual voice or create a new artificial voice is a system that "combines the two" functions into a single system, and not necessarily a system that combines a new created artificial voice with a mimicked individual voice. In fact, the Examiner's interpretation cannot be logically drawn from the language of column 6. Applicants' position in this regard is bolstered in the fact that nowhere else in the specification does it discuss or teach combining a mimicked individual voice with a "created new artificial voice." To do so would be illogical. Further, Applicants note that column 6 is part of the summary of the invention and thus, further details regarding the Examiner's approach should be found in the detailed description portion of the specification. However, Applicants can find no place where a "new" or "artificial" voice is combined with a "mimic" individual voice. Accordingly, Applicants respectfully submit that column 6, lines 60-62 do not provide the support that the Office Action implies for combining voices. Furthermore, as has been noted above, the Office Action mischaracterizes this step and when the appropriate scope of this limitation is properly assessed, Applicants submit that the feature of generating a new TTS voice by blending the first TTS voice and the second TTS voice and according to user selected voice characteristics are not taught in the portions of Gasper et al. cited in the Office Action.

Accordingly, Applicants submit that claim 1 is patentable over Gasper et al. and in condition for allowance.

Based on the foundation of claim 1 discussed above, Applicants further submit that many of the features of the dependent claims are not found in Gasper et al. contrary to what is asserted in the Office Action. For example, the Office Action asserts with regards to claim 2 that presenting a new TTS voice to the user for preview, receiving user-selected adjustments and presenting a revised TTS voice to the user for preview according to the user-selected adjustments is taught also in column 5, lines 50-64 and column 6, lines 58-65. Applicants respectfully submit

that there is no place in this portion of Gasper et al. in which such user interaction is taught. Accordingly, claim 2 is not anticipated by Gasper et al.

Claim 4 recites the method of claim 1, wherein the user-selected voice characteristic relates to mispronunciations. Applicants note that nothing in the columns 5 or 6 of Gasper et al. references mispronunciations. Accordingly, Applicants submit that Gasper et al. fail to teach this limitation and claim 4 is patentable and in condition for allowance.

Claim 6 recites the method of claim 5, wherein the prosodic characteristics are selected from a group comprising pitch contour, spectral envelope, volume contour and phone durations. The Office Action asserts also that columns 5 and 6 teach these limitations. However, Applicants cannot find reference to a spectral envelope, volume contour and phone durations in the discussion of prosodic features in the reference. Accordingly, Applicants submit that claim 6 is patentable and in condition for allowance. Similarly, claim 7 references a syllable accent, language accent and emotion. The reference does mention stress in column 5, line 56. Accordingly, Applicants have removed the word "stress" from claim 7 and thus, eliminating this as a member of the group. Accordingly, Applicants submit that the features listed in the group are not mentioned in Gasper et al.

Claim 8 recites wherein blending the first TTS voice and the second TTS voice further comprises extracting a prosodic characteristic from the LPC residual of the first TTS voice and the LPC residual of the second TTS voice and interpolating between the extracted prosodic characteristics. Applicants respectfully submit that the portion of columns 4 and 5 cited on page 4 of the Office Action and cited above in the present response, fail to teach anything regarding an interpolation between extracted prosodic characteristics. Accordingly, Applicants submit that claim 8 is patentable and in condition for allowance.

Similarly, claim 9 depends from claim 8 and further states wherein the interpolation of

the extracted pitches from the first TTS voice and the second TTS voice generates a new blended

pitch. Inasmuch as portions of Gasper et al. cited in the Office Action fail to teach anything

regarding interpolation of the extracted pitches, Applicants submit that claim 9 is also patentable

and in condition for allowance.

Claim 10 teaches a method of generating a synthetic voice which includes receiving a

user selection of TTS voice and a voice characteristic and presenting a user with a new TTS

voice comprising the selected TTS voice blended with at least one other TTS voice to achieve

the selected voice characteristic. Applicants respectfully submit that the portions of Gasper et al.

cited on page 5 are discussed above and the above argument applies to the limitations of claim

10. For example, Applicants submit that the reference fails to teach presenting the user with a

new TTS voice that includes a selected TTS voice that is blended with at least one other TTS

voice to achieve the selected voice characteristics. Accordingly, Applicants submit that claim 10

is patentable. Claims 11-17 and 19-20 each depend from claim 10 and recite further limitations

therefrom. Applicants note that claim 19 is amended to be dependent from claim 16 rather than

cancelled claim 18. Applicants respectfully submit that similar limitations discussed above

relative to claim 1 also apply to the claims dependent from claim 10. Accordingly, Applicants

submit that claim 10 and its dependent claims are patentable and in condition for allowance.
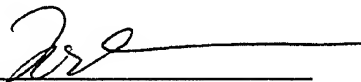
Applicants also submit that claims 21-35 are patentable for the same reasons set forth

above.

CONCLUSION

Having addressed all rejections and objections, Applicants respectfully submit that the

subject application is in condition for allowance and a Notice to that effect is earnestly solicited.

If necessary, the Commissioner for Patents is authorized to charge or credit the **Novak, Druce &**

Quigg, LLP, Account No. 14-1437 for any <u>deficiency or overpayment</u>.

Respectfully submitted,

Date: November 26, 2007   By: _____

<u>Correspondence Address:</u>   Thomas M. Isaacson
Thomas A. Restaino
Reg. No. 33,444   Attorney for Applicants
AT&T Corp.   Reg. No. 44,166
Room 2A-207   Phone: 410-286-9405
One AT&T Way   Fax No.: 410-510-1433
Bedminster, NJ 07921